

# A Generalization of a Bin-Based Modification to the Stochastic Simulation Algorithm

David Collins

*Dept. of Mathematics and Statistics, University of Victoria  
PO Box 1700 STN CSC, Victoria, B.C. Canada V8W 2Y2*

November 24, 2015

## Abstract

In a 1996 paper, Seeßelberg, Trautmann and Thorn [12] modified Gillespie's [6] Monte Carlo algorithm which is used to stochastically simulate the collision and coalescence process. Their modification reduces the storage requirements of the simulation by several orders of magnitude. However, their modification creates unphysical and potentially fatal conditions when used with common initial distributions. We identify those conditions and propose a solution to maintain physically real conditions for all state variables during the evolution of any initial distribution.

Keywords: stochastic simulation, cloud microphysics, collision, coalescence, droplet spectrum

## 1 Introduction

Gillespie [6] presented a Monte Carlo algorithm that stochastically simulates the collision and coalescence of liquid droplets. His algorithm required the recording of information about  $N$  droplets at each time step which demands impractically large storage requirements. Seeßelberg et. al. [12] reduced the required storage by discretizing the droplet size spectrum such that  $N$  droplets are contained in  $n$  bins. Each bin contains an aggregate of droplets. Thus, bin  $i$  contains an aggregate mass  $m_i$  and a aggregate droplet number  $N_i$ . The modification of Seeßelberg et. al. requires the storage of  $2n$  variables (mass and number for each bin) rather than the masses of  $N$  droplets. Storage requirements are significantly reduced since  $N \gg 2n$  by  $\sim O(6)$ .

Gillespie accounted for every drop in his simulation. This demanded that droplet numbers be non-negative integers. He described his simulation as encompassing an entire cloud which suggests the interpretation of the droplet mass as being represented by the 'continuous model' of coalescence where the evolution of the droplet distribution tracks the mass of every droplet [5]. His exact method was theoretical. Although it did provide algorithms to execute the theory, the computational requirements of the algorithms were not quantifiably scaled to actual clouds as was his theory. The computational requirements of his exact method when scaled to simulate entire clouds cannot be met by the most powerful computers of today.

However, a stochastic simulation can model droplet collisions in a subset of a cloud, and the results of such a simulation could then be scaled over the entire cloud. The simulated volume could be considered a spatial average. Assuming ergodicity as shown by Pope [17] an ensemble of

simulations, and subsequent spatial average, suggested by Gillespie [5] would satisfy the expected variation over a homogeneous cloud, but would not account for entrainment or detrainment. Such an interpretation would be consistent with ‘quasi-stochastic’ or ‘pure stochastic’ models of coalescence where the spectral evolution tracks the number of drops of each mass or the probability of any number of drops of each mass, respectively, as explained in one of Gillespie’s 1975 papers [5].

Seeßelberg and collaborators apparently followed Gillespie’s precedent and developed their bin-based modification using the ‘continuous model’ of coalescence which assigned non-negative integers to aggregate droplet numbers  $N_i$  for each bin. They did not specify that aggregate droplet numbers  $N_i$  in bin  $i$  would be restricted to integer values. However, non-integer values can create unphysical droplet values for bins with  $1 < N_i < 2$  as is shown in Section 3.1. A simulation volume containing non-integer values of number concentrations would be consistent with ‘quasi-stochastic’ or ‘pure stochastic’ models of coalescence [5]. They used an initial gamma distribution. To construct such a distribution using integer values for droplet numbers requires a crude approximation of the gamma distribution. An approximation of, otherwise smooth, initial distributions may create bias in the evolved distribution due to non-uniform adjustments to intra-bin number concentrations. Whether they made such crude approximations for initial distributions or ran enough simulations that a reasonably sized subset was successful is unclear. Only the latter option creates non-physical state spaces as described in Section 3.1.

The Monte Carlo algorithm and the subsequent bin-based modification are methods used to simulate an exact evolution of moments of a probability distribution that may otherwise be evolved by the kinetic collection equation. A review of the literature identifies the Seeßelberg et. al. 1996 paper as a reference when validating the use of bin-based moment methods such as Bott’s Linear Flux Method [8, 10], when identifying a Monte Carlo method [3, 7, 14], or when quantifying a coalescence efficiency [11, 14]. In 1999 Trautmann, a coauthor of the 1996 article, expanded upon that paper, with other collaborators, to include an exponential in-bin distribution and log-radius coordinates [15]. In 1999 Tzivion and collaborators critiqued the work of Seeßelberg et. al. by noting that it did not include sensitivity tests w.r.t. number of bins, kernel used, the time step, or initial conditions [16]. Other than the 1999 article by Trautmann et. al., a reproduction of the bin-based modification to the stochastic simulation has not been found in the literature by this author.

Generalizing the bin-based modification of Gillespie’s stochastic simulation to include positive non-integer droplet numbers is consistent with a ‘quasi-stochastic’ or ‘pure stochastic’ model of coalescence [5]. When a stochastic algorithm uses non-integers to quantify droplet numbers in simulations of a subset of a physical cloud, the evolved droplet size information can be interpreted as the number of drops of each mass or the probability of any number of drops of each mass. In either case by the ergodicity shown by Pope [17], an ensemble of simulations of the generalized bin-based, modified stochastic algorithm gives a mean and variance of droplet masses that can be interpreted to represent a homogeneous portion of a cloud.

These first and second moment statistics can be used to verify the results of detailed spectral methods. A discrete spectral method is, in turn, used to validate bulk parameterizations [9, 13]. Accurate bulk parameterizations are used by cloud resolving models to determine the onset of precipitation and the amount of radar reflectivity [4, 13]. To make this necessary process more widely accessible we present a further modification that eliminates conditions leading to unphysical state spaces.

The discretization process introduces physical requirements on the modified evolution algorithm

that were not present when individual droplets were selected by Gillespie’s Monte Carlo Method. The mean mass within a bin is defined (for bin  $i$ ) by  $\bar{x}_i = m_i/N_i$ . For the rest of this paper the phrase ‘mean mass’ will refer to the mean mass within a bin. Each collision in a simulation of the modified Gillespie algorithm involves reducing the droplet number of each selected source bin by unity and choosing, for removal, a droplet mass within the source bin. The droplet mass must be chosen so that the remaining mass and remaining number yields a mean mass that is within the bin boundaries. In Section 2 the droplet selection process of the modified Monte Carlo algorithm by Seeβelberg and collaborators is briefly reviewed. In Section 3 the existence and prevalence of conditions leading to unphysical state spaces are presented ,and an example is given. Section 4 details the modification to the droplet selection process used to eliminate unphysical conditions.

## 2 Current Intra-Bin Droplet Selection Process

The algorithm by Seeβelberg et. al. includes a process to select a subinterval within the bin from which to choose a droplet mass. The subinterval is given in their paper as Equation (9). Seeβelberg et. al. assumed a uniform distribution of mass density within this subinterval and randomly selected a droplet. In Sections 3 and 4 we maintain that assumption. The interval  $I_i$  is constructed to be symmetric about the mean mass  $\bar{x}_i$ . The masses of droplets at the left and right-hand bin boundaries are denoted as  $x_{i-1/2}$  and  $x_{i+1/2}$ , respectively. Letting  $\Delta m_i \equiv \min(x_{i+1/2} - \bar{x}_i, \bar{x}_i - x_{i-1/2})$ , the bounds of  $I_i$  are defined by

$$d_{i-1/2} = \bar{x}_i - \Delta m_i, \quad \text{and} \quad d_{i+1/2} = \Delta m_i + \bar{x}_i \quad [12]$$

$$I_i = [d_{i-1/2}, d_{i+1/2}] \subseteq [x_{i-1/2}, x_{i+1/2}]$$

## 3 Unphysical Conditions

The existence if the conditions that facilitate the unphysical state space is discussed. An example which arises from the initial distribution used by Seeβelberg et. al. is detailed, and the prevalence of the condition is noted.

### 3.1 Existence

The droplet selection process given by Seeβelberg et. al. fails to adequately restrict the interval  $I_i$  when  $1 < N_i < 2$ . An example of the conditions predicating this problem are present in the initial droplet size distribution used by Seeβelberg et. al., and commonly used by others [1–4]. This initial gamma distribution is given in the paper by Seeβelberg et. al. as their Equation (19). When for example bin  $i$  containing  $N_i \in (1, 2)$  droplets is selected as the bin containing a pre-collision droplet for the next collision, the removal of too much or too little mass is possible and would result in a post-collision mean mass for bin  $i$  that is either less than the mass of the smaller bin boundary, or greater than mass of the larger bin boundary, respectively.

The sketch in Figure (1) illustrates conditions that can lead to a mean mass outside of the bin after the collision. The spectrum is described by the mass of a droplet where  $\hat{g}_i$  is the total mass in bin  $i$ . The two larger (red) arcs are the bounds of interval  $I_i$ . A sub-interval bounded by  $\hat{g}_i$  and  $x_{i+1/2}$

lies within the interval bounded by the larger (red) arcs. Therefore, when  $N_i$  is slightly greater than unity, say  $N_i = 1 + \alpha$  it is, (i) possible to select an amount of mass to be removed that is greater than the total amount of mass in bin  $i$ , and (ii) possible to select an amount of mass that is less than  $\hat{g}_i$  but will result in the post-collision mean mass less than  $x_{i-1/2}$  as further detailed in the following example.

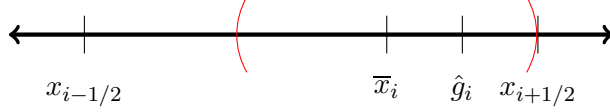


Figure 1: A schematic of one bin on the droplet size spectrum showing the bin boundaries for bin  $i$ , mean intra-bin mass  $\bar{x}_i$ , and the aggregate mass contained in bin  $i$ ,  $\hat{g}_i$ . The larger (red) arcs are the boundaries of the interval within which a droplet is chosen for the next collision. The algorithm to determine these boundaries are by Seeβelberg et. al. (1996) and briefly explained in Section 2.

### 3.2 Example

The initial conditions presented in Seeβelberg et. al. create conditions in bin #17 of their spectral discretization scheme that potentially lead to a post collision mean mass outside of the bin boundaries even when the mass removed is restricted to an amount less than the mass present. The scaling of the initial distribution by 1500 as reported in their article only serves to shift the bin with  $1 < N_j < 2$  downstream. Bin #17 of their unscaled distribution contains  $2.046 \times 10^{-08} \text{g}$  and 1.089 droplets. When  $1.893 \times 10^{-08} \text{g}$  are removed then the remaining mass  $x_{17,r}$  and remaining droplet number 0.089 gives a mean mass equal to the mass of the smaller bin boundary. Therefore, when a droplet of mass  $x_{17,r}$  is removed from bin #17 and  $x_{17,r} > 1.893 \times 10^{-08} \text{g}$ , the post collision mean mass will be greater than the upper bound of the bin #17.

This flaw can be expressed as  $\hat{x}_i < x_{i,r} < \hat{g}_i$  where  $\hat{x}_i$  is the maximum mass that can be removed while keeping the post collision mean mass inside of the bin. Having a mean mass outside of a bin that contains a fraction of a droplet such as 0.089 may not seem critical, but the addition of a droplet to bin  $i$ , due to a collision with a smaller droplet, will not guarantee that the mean mass will return to the inside of the bin.

### 3.3 Prevalence

The prevalent use of a gamma function as an initial distribution requires that bins with  $1 < N_i < 2$  will typically be present. The persistence of a practical maximum droplet size (whether through precipitation or breakup) guarantees the existence of the condition in some bin at each time step. A distribution which develops a bimodal character can have multiple bin with potential fatalities. The stochastic nature of the algorithm facilitates a random selection of a source bin, and the possibility that a bin with a critically low droplet number be chosen.

Several simulations have been run without the refinement presented in this paper, and many have produced a mean mass external to the bin. These external mean mass occurrences have occurred in multiple bins and at various time steps. Successfully completing a simulation without a fatal error is possible. However, such results may be produced after the existence of an unphysical state

space at some time step during the simulation which may call into question the validity of the simulation.

## 4 A Refinement to the Droplet Selection Process

A sub-interval  $\hat{I}_i$  is constructed that is dependent on the bin boundaries ( $x_{i-1/2}$  and  $x_{i+1/2}$ ):  $\hat{I}_i \subseteq I_i$  such that after the simulation of a collision is complete the mean mass  $\bar{x}_i$  will remain between the bin boundaries. The upper bound of  $\hat{I}_i$  is shown in Figure (2) and subsequently developed herein.

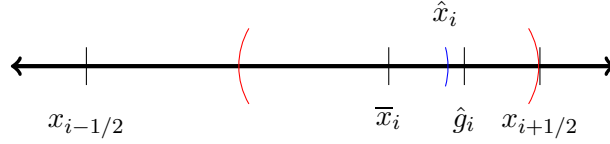


Figure 2: A schematic of one bin on the droplet size spectrum showing the same as in Figure (1) with the addition of a smaller (blue) arc  $\hat{x}_i$  that represents the upper bound of a more restrict interval. The calculations used to create this more restrictive boundary are described in the surrounding text.

The smaller (blue) arc represents the maximum amount of mass that can be removed such that  $\hat{g}_i - \hat{x}_i$  is the amount of mass remaining after the next collision involving bin  $i$  and  $(\hat{g}_i - \hat{x}_i) / \alpha = \bar{x}_{i,a} = x_{i-1/2}$  where  $\bar{x}_{i,a}$  is the mean mass after the collision. The following procedure is used to calculate  $\hat{x}_i$  as a function of  $x_{i-1/2}$  and also to calculate a more restrictive lower bound as a function of  $x_{i+1/2}$ .

The following procedure will give one or more bounds of the more restrictive interval that ensures the remaining mean intra-bin mass will be within the bin boundaries:

$$\begin{aligned} e_{i-1/2} &= L_i - x_{i+1/2}(N_i - 1) & e_{i+1/2} &= \hat{x}_i = L_i - x_{i-1/2}(N_i - 1) \\ \bar{d}_{i-1/2} &= \max(d_{i-1/2}, e_{i-1/2}), & \bar{d}_{i+1/2} &= \min(d_{i+1/2}, e_{i+1/2}) \\ \bar{I}_i &= [\bar{d}_{i-1/2}, \bar{d}_{i+1/2}] \subseteq I_i \subseteq [x_{i-1/2}, x_{i+1/2}] \end{aligned}$$

$L_i$  and  $N_i$  are the liquid water content and number concentration in bin  $i$  prior to the next collision involving bin  $i$ .

In the example from Section 3.2,  $\bar{d}_{i+1/2}$  is a function of the lower bin boundary and is shown in Figure (3).

Notice that the interval bounded by the left-most of the larger (red) arcs and  $\hat{x}_i$  is not symmetric about  $\bar{x}_i$ . An interval symmetric about  $\bar{x}_i$  is needed to avoid bias. Therefore, a final step in the spirit of the algorithm by Seeßelberg et. al. (1996) is used to produce the sub-interval  $\hat{I}_i$  bounded by the two smaller (blue) arcs:

$$\hat{x}'_i = \bar{x}_i - (\hat{x}_i - \bar{x}_i)$$

We now have an interval  $\hat{I}_i = [\hat{x}'_i, \hat{x}_i]$ , symmetric about  $\bar{x}_i$ , from which to draw the pre-collision droplet which guarantees that the remaining mass and remaining number concentration will yield a post-collision mean intra-bin mass that is inside of the bin boundaries:

$$\hat{I}_i = [\hat{x}'_i, \hat{x}_i] \subseteq \bar{I}_i \subseteq I_i \subseteq [x_{i-1/2}, x_{i+1/2}]$$

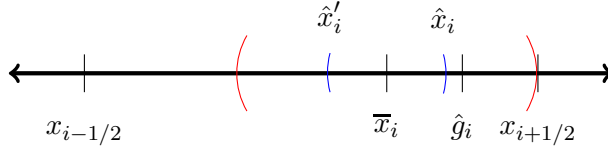


Figure 3: A schematic of one bin on the droplet size spectrum showing the same as in Figure (2) with the addition of a second smaller (blue) arc  $\hat{x}'_i$  that represents the lower bound of a more restrictive interval. The calculations used to create the lower bound of this more restrictive interval are in the same spirit as that originally used by Seeßelberg et. al. (1996).

## 5 Conclusion

The discretization of the droplet spectrum by Seeßelberg (1996) and collaborators reduced the computational expense of stochastically simulating the collision and coalescence process over the expense of a stochastic simulation that records information about every droplet as given by Gillespie (1975). That modification produce the additional requirement of ensuring that the ratio of intra-bin mass and intra-bin number lie between the two values on the edges of the bin. This requirement is intrinsically met by the algorithm presented by Seeßelberg et. al. except for the cases when the post-collision intra-bin number is  $0 < N_i < 1$ . They restrict the interval in the source bins for the purpose of eliminating droplet selection bias. However, their selection interval does not guarantee that the post-collision mean mass lies within the bin. Situations resulting in such a post collision number concentration has been shown to occur in the commonly used gamma initial distribution, and the condition will persist as the mode traverses the droplet spectrum.

These situations are addressed by further restricting the intra-bin interval contained in the source bins upon which is constructed a uniform distribution. The additional restrictions on the domain from which the mass of a source droplet can be selected are functions of bin boundaries. The new domain is then made symmetric about the intra-bin mean mass to eliminate selection bias. The restricted intra-bin domain allows for a generalization of initial conditions that include continuous (i.e. non-integer) values of number concentration. There is no need to modify an initial distribution that contains non-integer values of number concentration. Nor is there a need to prevent collisions of droplets in bins containing a number between 1 and 2.

## References

- [1] Edwin Berry. Cloud droplet growth by collection. *J. Atmos. Sci.*, 24:688–701, 1967.
- [2] Edwin Berry and Richard Reinhardt. An analysis of cloud droplet growth by collection: Part i. double distributions. *J. Atmos. Sci.*, 31:1814–1824, 1974.
- [3] Andreas Bott. A flux method for the numerical solution of the stochastic collection equation. *J. Atmos. Sci.*, 55:2284–2293, 1998.
- [4] Charmaine Franklin. A warm rain microphysics parameterization that includes the effects of turbulence. *J. Atmos. Sci.*, 65:1795–1816, 2008.

- [5] Daniel T. Gillespie. Three models for the coalescence growth of cloud droplets. *J. Atmos. Sci.*, 32:600–607, 1975a.
- [6] Daniel T. Gillespie. An exact method for numerically simulating the stochastic coalescence process in a cloud. *J. Atmos. Sci.*, 32:1977–1989, 1975b.
- [7] A. Khain, V. Arkhipov, M. Pinsky, Y. Feldman, and Y. Ryabov. Rain enhancement and fog elimination by seeding with charged droplets. *J. Appl. Meteor.*, 43:1513–1529, 2004.
- [8] A. Khain, M. Ovtchinnikov, M. Pinsky, A. Pokrovsky, and H. Krugliak. Notes on the state-of-the-art numerical modelling of cloud microphysics. *Atmos. Res.*, 55:159–224, 2000.
- [9] Marat Khairoutdinov and Yefim Kogan. A new cloud physics parameterization in a large eddy simulation model of marine stratocumulus. *Mon. Wea. Rev.*, 128:229–243, 2000.
- [10] M. B. Pinsky and A. Khain. Effects of in-cloud nucleation and turbulence on droplet spectrum formation in cumulus clouds. *Q. J. R. Meteorol. Soc.*, 128:501–553, 2002.
- [11] M. B. Pinsky, A. Khain, D. Rosenfeld, and A. Pokrovsky. Comparison of collision velocity differences of drops and graupel particles in a very turbulent cloud. *Atmos. Res.*, 49:99–113, 1998.
- [12] M Seesselberg, T Trautmann, R McGraw, and M Thorn. Stochastic simulations as a benchmark for mathematical methods solving the coalescence equation. *Atmos. Res.*, 40:33–48, 1996.
- [13] Axel Seifert and Klaus D. Beheng. A double moment parameterization for simulating auto-conversion, accretion, and self-collection. *J. Atmos. Sci.*, 59-60:265–281, 2001.
- [14] Shima Shin-ichiro, Kanya Kusano, Akio Kawano, Tooru Sugiyama, and Shintaro Kawahara. Super-droplet method for the numerical simulation of clouds and precipitation: A particle-based microphysics model coupled with non-hydro-static model. *Atmospheric and Oceanic Physics*, 135:1307–1320, 2009.
- [15] T. Trautmann and C. Wanner. A fast and efficient modified sectional method for simulating multicomponent collisional kinetics. *Atmospheric Environment*, 33:1631–1640, 1999.
- [16] S. Tzivion, T. Reisin, and Z. Levin. A numerical solution of the kinetic collection equation using high spectral resolution: A proposed reference. *J. Comp. Physics*, 148:527–544, 1999.
- [17] John C. Wyngaard. *Turbulence in the Atmosphere*. Cambridge University Press, New York, 2010.